

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Appln. No: To Be Assigned  
Applicant: Keiko Morii et al.  
Filed: Herewith  
Title: METHOD OF SPEAKER NORMALIZATION FOR SPEECH RECOGNITION  
USING FREQUENCY CONVERSION AND SPEECH RECOGNITION  
APPARATUS APPLYING THE PRECEDING METHOD  
TC/A.U.:  
Examiner:  
Docket No.: MAT-8466US

**CLAIM TO RIGHT OF PRIORITY**

Mail Stop Patent Application  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

Pursuant to 35 U.S.C. § 119, Applicant(s) hereby claim the benefit of prior Japanese Patent Application No. 2002-277022, filed September 24, 2002.

A certified copy of the above-referenced application is enclosed.

Respectfully submitted,

PatnerPrestia

Lawrence E. Ashery, Reg. No. 34,515  
Attorney for Applicants

LEA/kc

Enclosure: Certified Copy of Patent Application No. JP 2002-277022

Dated: September 24, 2003

P.O. Box 980  
Valley Forge, PA 19482-0980  
(610) 407-0700

The Commissioner for Patents is hereby  
authorized to charge payment to Deposit  
Account No. 18-0350 of any fees associated  
with this communication.

EXPRESS MAIL

Mailing Label Number:  
Date of Deposit:

EV 351885936 US  
September 24, 2003

I hereby certify that this paper and fee are being deposited, under 37 C.F.R. § 1.10 and with sufficient postage, using the "Express Mail Post Office to Addressee" service of the United States Postal Service on the date indicated above and that the deposit is addressed to the Mail Stop Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Kathleen Libby  
Kathleen Libby

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日

Date of Application:

2002年 9月24日

出願番号

Application Number:

特願2002-277022

[ST.10/C]:

[JP2002-277022]

出願人

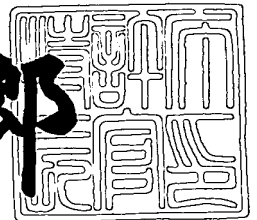
Applicant(s):

松下電器産業株式会社

2003年 6月 2日

特許庁長官  
Commissioner,  
Japan Patent Office

太田信一郎



出証番号 出証特2003-3042683

【書類名】 特許願

【整理番号】 2931040044

【提出日】 平成14年 9月24日

【あて先】 特許庁長官殿

【国際特許分類】 G10L 15/10

【発明者】

    【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式  
                                会社内

    【氏名】 森井 景子

【発明者】

    【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式  
                                会社内

    【氏名】 中藤 良久

【発明者】

    【住所又は居所】 大阪府門真市大字門真 1 0 0 6 番地 松下電器産業株式  
                                会社内

    【氏名】 桑野 裕康

【特許出願人】

    【識別番号】 000005821

    【氏名又は名称】 松下電器産業株式会社

【代理人】

    【識別番号】 100097445

    【弁理士】

    【氏名又は名称】 岩橋 文雄

【選任した代理人】

    【識別番号】 100103355

    【弁理士】

    【氏名又は名称】 坂口 智康

【選任した代理人】

【識別番号】 100109667

【弁理士】

【氏名又は名称】 内藤 浩樹

【手数料の表示】

【予納台帳番号】 011305

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9809938

【書類名】 明細書

【発明の名称】 話者正規化方法及びそれを用いた音声認識装置

【特許請求の範囲】

【請求項 1】 入力音声のフレーム毎の音響特徴量を抽出する特徴量抽出ステップと、  
前記音響特徴量をあらかじめ定められた周波数変換係数を用いて周波数変換する周波数変換ステップと、  
前記周波数変換した変換後特徴量及び標準音韻モデルを用いてフレーム毎に類似度あるいは距離を比較して、最尤の音韻及び最適周波数変換係数を算出するステップと、  
前記最尤の音韻及び前記最適周波数変換係数を用いて、前記入力音声を正規化する周波数変換係数を選択するステップと、  
前記選択された周波数変換係数を用いて周波数変換をするステップと  
を有する話者正規化方法。

【請求項 2】 前記選択は、前記入力音声を正規化する周波数変換係数を、前記最適周波数変換係数となった頻度の最も多い周波数変換係数に決定する請求項 1 に記載の話者正規化方法。

【請求項 3】 前記選択するステップは、複数の前記フレームでの最尤の音韻に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を当該最尤の音韻に対する入力音声最適変換係数と決定するステップと、  
前記フレーム毎の前記最尤の音韻に対する前記入力音声最適変換係数を当該フレームの前記入力音声を正規化する周波数変換係数に決定するステップと  
からなる請求項 1 に記載の話者正規化方法。

【請求項 4】 前記音響特徴量と前記標準音韻モデルを用いて、フレーム毎に各音韻同士の類似度あるいは距離の比を重みとして算出するステップと、  
前記入力音声を正規化する前記周波数変換後に算出する類似度あるいは距離に、  
前記重みを掛け合わせるステップと  
を更に有する請求項 1 に記載の話者正規化方法。

【請求項 5】 前記選択するステップは、複数の前記フレームでの最尤の音韻

に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を、当該最尤の音韻に対する入力音声最適変換係数と決定するステップと、  
前記音韻毎の前記重みを、前記フレーム毎の前記最適周波数変換係数の重みに変換するステップと、  
前記重みが0でない前記最適周波数変換係数を、当該フレームの前記入力音声を正規化する周波数軸変換係数に決定するステップと  
からなる請求項4に記載の話者正規化方法。

【請求項6】 前記類似度あるいは前記距離の比較には少なくとも母音を使用することを特徴とする請求項1乃至5に記載の話者正規化方法。

【請求項7】 前記類似度あるいは前記距離の比較は母音のみを使用することを特徴とする請求項1乃至5に記載の話者正規化方法。

【請求項8】 入力音声のフレーム毎の音響特徴量を抽出する特徴量抽出部と、  
前記音響特徴量をあらかじめ定められた周波数変換係数を用いて周波数変換する周波数変換部と、  
前記周波数変換した変換後特徴量及び標準音韻モデルを用いてフレーム毎に類似度あるいは前記距離を比較して、最尤の音韻及び最適周波数変換係数を算出する音響特徴量比較部と、  
前記最尤の音韻及び前記最適周波数変換係数を用いて、前記入力音声を正規化する周波数変換係数を選択する変換条件判定部と、  
前記周波数変換部からの変換後特徴量及び認識対象標準音響モデルとを比較して音声を認識する音声認識処理部とを備え、  
前記音声認識処理部が、前記選択された周波数変換係数を用いて、再度周波数変換された変換後特徴量を用いて前記認識を行うことを特徴とする音声認識装置。

【請求項9】 前記変換条件判定部が、前記入力音声を正規化する周波数変換係数を、前記最適周波数変換係数となった頻度の最も多い周波数変換係数として選択する請求項8に記載の音声認識装置。

【請求項10】 前記変換条件判定部が、複数の前記フレームでの最尤の音韻に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を当該最尤

の音韻に対する入力音声最適変換係数と決定し、更に前記フレーム毎の前記最尤の音韻に対する前記入力音声最適変換係数を当該フレームの前記入力音声を正規化する周波数変換係数に決定することを特徴とする請求項 8 に記載の音声認識装置。

【請求項 1 1】 前記音韻特徴量及び前記標準音韻モデルを用いて、フレーム毎に各音韻同士の類似度あるいは距離の比を重みとして算出する重み算出部を更に備え、  
前記音声認識部が前記入力音声を正規化する前記周波数変換後に算出する類似度あるいは距離に、前記重みを掛け合わせたものを使用して認識することを特徴とする請求項 8 に記載の音声認識装置。

【請求項 1 2】 前記変換条件判定部が、複数の前記フレームでの最尤の音韻に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を、当該最尤の音韻に対する入力音声最適変換係数と決定し、更に前記音韻毎の前記重みを、前記フレーム毎の前記最適周波数変換係数の重みに変換し、前記重みが 0 でない前記最適周波数変換係数を、当該フレームの前記入力音声を正規化する周波数軸変換係数に決定することを特徴とする請求項 1 1 に記載の音声認識装置。

【請求項 1 3】 前記類似度あるいは前記距離の比較は少なくとも母音を使用することを特徴とする請求項 8 乃至 1 2 に記載の音声認識装置。

【請求項 1 4】 前記類似度あるいは前記距離の比較は母音のみを使用することを特徴とする請求項 8 乃至 1 2 に記載の音声認識装置。

#### 【発明の詳細な説明】

##### 【0 0 0 1】

##### 【発明の属する技術分野】

本発明は音響特徴量を用いた音声の個人差を補正する話者正規化方法及び、それを用いた音声認識装置に関する。

##### 【0 0 0 2】

##### 【従来技術】

従来より、この話者正規化方法を用いた音声認識装置としては、特許文献 1 に記載されているようなものがあった。図 1 2 は、特許文献 1 の従来音声認識方

法を示す流れ図である。

【0003】

まず、A/D変換を行ってデジタル化された音声を入力信号として、LPCケプストラム係数等の特徴量を抽出し（ステップS10）、次に、発声者の声道長の個人差に起因する影響を正規化するために、LPCケプストラム等の特徴量に周波数軸上の変換を施す（ステップS30）。

【0004】

次に、周波数軸上の変換を施された入力音声の特徴量と予め複数話者から学習した音響モデル特徴量とのマッチングを行う（ステップS50）。その後、ステップS50において算出された認識結果をもとに入力発声を教師信号として最適な変換係数を求める（ステップS60）。次に、話者や音韻によるばらつきを吸収するため変換係数平滑化を行い、新たな周波数変換係数を更新する（ステップS70）。

【0005】

【特許文献1】

特開2001-255886号公報（第4-8頁、第2図）

【0006】

【発明が解決しようとする課題】

しかしながら、このような従来の方法では、無声音と有声音の区間検出を行った後に、情報に同期して周波数変換を行う必要がある上に、認識対象語辞書が話者正規化を行う際にも必要であるという課題を有していた。

【0007】

本発明は、このような従来の課題を解決するものであり、入力音声の個人差を補正して、認識対象語辞書を使用することなく話者正規化を行うことで音声認識性能を向上せしめることを目的とする。

【0008】

【課題を解決するための手段】

上記従来の課題を解決する本発明の話者正規化方法は、入力音声のフレーム毎の音響特徴量を抽出する特徴量抽出ステップと、この音響特徴量をあらかじめ定



められた周波数変換係数を用いて周波数変換する周波数変換ステップと、この周波数変換した変換後特徴量と標準音韻モデルを用いてフレーム毎に類似度あるいは距離を比較して、最尤の音韻と最適周波数変換係数とを算出するステップと、最尤の音韻と最適周波数変換係数を用いて、入力音声を正規化する周波数変換係数を選択するステップと、この選択された周波数変換係数を用いて周波数変換をするステップとを有する。

## 【 0 0 0 9 】

また、本発明の音声認識装置は、入力音声のフレーム毎の音響特徴量を抽出する特徴量抽出部と、この音響特徴量をあらかじめ定められた周波数変換係数を用いて周波数変換する周波数変換部と、この周波数変換した変換後特徴量と標準音韻モデルを用いてフレーム毎に類似度あるいは距離を比較して、最尤の音韻と最適周波数変換係数とを算出する音響特徴量比較部、最尤の音韻と最適周波数変換係数を用いて、入力音声を正規化する周波数変換係数を選択する変換条件判定部と、周波数変換部からの変換後特徴量と認識対象標準音響モデルとを用いて音声を認識する音声認識処理部とを備え、音声認識処理部が、選択された周波数変換係数を用いて、再度周波数変換された変換後特徴量を用いて認識を行うことを特徴としている。

## 【 0 0 1 0 】

これにより、入力音声を標準話者の音響特徴量に合わせるように正規化することによって入力音声の話者による差が正規化され、認識対象語辞書を使用することなく認識性能が向上することができる。

## 【 0 0 1 1 】

## 【発明の実施の形態】

本発明の話者正規化方法は、入力音声のフレーム毎の音響特徴量を抽出する特徴量抽出ステップと、この音響特徴量をあらかじめ定められた周波数変換係数を用いて周波数変換する周波数変換ステップと、この周波数変換した変換後特徴量と標準音韻モデルを用いてフレーム毎に類似度あるいは距離を比較して、最尤の音韻と最適周波数変換係数とを算出するステップと、最尤の音韻と最適周波数変換係数を用いて、入力音声を正規化する周波数変換係数を選択するステップと、

この選択された周波数変換係数を用いて周波数変換をするステップとを有する。

【0012】

これによって、入力音声を標準話者の音響特徴量に合わせるように正規化することによって入力音声の話者による差が正規化され、認識性能を向上させることができる。

【0013】

また、本発明の話者正規化方法の選択は、入力音声を正規化する周波数変換係数を、最適周波数変換係数となった頻度の最も多い周波数変換係数に決定する。

【0014】

これによって、簡単な演算により話者正規化を行うことができる。

【0015】

また、本発明の話者正規化方法の選択するステップは、複数のフレームでの最尤の音韻に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を当該最尤の音韻に対する入力音声最適変換係数と決定するステップと、フレーム毎の最尤の音韻に対する入力音声最適変換係数を当該フレームの入力音声を正規化する周波数変換係数に決定するステップとからなっている。

【0016】

これによって、各入力フレームの周波数変換を行う変換係数が1つ選定されるが、入力フレーム毎に選定される変換係数は異なるため、1フレーム毎により細かく話者正規化を行うことが可能になり、いかなる音声に対してもこの話者正規化方法を適用することができる。

【0017】

また、本発明の話者正規化方法は、音響特徴量と標準音韻モデルを用いて、フレーム毎に各音韻同士の類似度あるいは距離の比を重みとして算出するステップと、入力音声を正規化する前記周波数変換後に算出する類似度あるいは距離に、前記重みを掛け合わせるステップとを更に有している。

【0018】

これによって、各入力フレームのスペクトルを周波数変換する変換係数は複数選定され、重み付き合算処理が行われ、かつ入力フレーム毎に重みの組値が異なる

る。このため、1フレーム毎により精度良く話者正規化を行うことが可能になり、いかなる音声に対してもこの話者正規化方法を適用することができる。

【0019】

また、本発明の話者正規化方法の選択するステップは、複数の前記フレームでの最尤の音韻に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を当該最尤の音韻に対する入力音声最適変換係数と決定するステップと、音韻毎の重みを、フレーム毎の最適周波数変換係数の重みに変換するステップと、重みが0でない最適周波数変換係数を、当該フレームの入力音声を正規化する周波数軸変換係数に決定するステップとからなっている。

【0020】

これによって、各入力フレームのスペクトルを周波数変換する変換係数は複数選定され、重み付き合算処理が行われ、かつ入力フレーム毎に重みの組値が異なる。このため、1フレーム毎により精度良く話者正規化を行うことが可能になり、いかなる音声に対してもこの話者正規化方法を適用することができる。

【0021】

また、本発明の話者正規化方法は、類似度あるいは距離の比較は少なくとも母音を使用することを特徴としている。

【0022】

これによって、周波数変換を行う対象として最も信頼性の高い母音のみの情報から全区間の最適変換係数を決定することになるため、決定された最適変換係数の信頼度もより高いものにすることができる。

【0023】

また、本発明の話者正規化方法は、類似度あるいは距離の比較は母音のみを使用することを特徴としている。

【0024】

これによって、音声認識性能が劣化することなく、標準音韻モデルのデータ量を少なく抑えることができる。

【0025】

本発明の音声認識装置は、入力音声のフレーム毎の音響特徴量を抽出する特徴

量抽出部と、この音韻特徴量をあらかじめ定められた周波数変換係数を用いて周波数変換する周波数変換部と、この周波数変換した変換後特徴量と標準音韻モデルを用いてフレーム毎に類似度あるいは距離を比較して、最尤の音韻と最適周波数変換係数とを算出する音響特徴量比較部、最尤の音韻と最適周波数変換係数を用いて、入力音声を正規化する周波数変換係数を選択する変換条件判定部と、周波数変換部からの変換後特徴量と認識対象標準音響モデルとを用いて音声を認識する音声認識処理部とを備え、音声認識処理部が、選択された周波数変換係数を用いて、再度周波数変換された変換後特徴量を用いて音声の認識を行うことを特徴としている。

## 【 0 0 2 6 】

これによって、入力音声を標準話者の音韻モデルに合わせるように正規化することによって入力音声の話者による差が正規化され、認識性能を向上させることができる。

## 【 0 0 2 7 】

また、本発明の音声認識装置の変換条件判定部は、入力音声を正規化する周波数変換係数を、最適周波数変換係数となった頻度の最も多い周波数変換係数として選択する。

## 【 0 0 2 8 】

これによって、簡単な演算により話者正規化を行うことができる。

## 【 0 0 2 9 】

また、本発明の音声認識装置の変換条件判定部は、複数のフレームでの最尤の音韻に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を当該最尤の音韻に対する入力音声最適変換係数と決定し、更にフレーム毎の最尤の音韻に対する入力音声最適変換係数を当該フレームの入力音声を正規化する周波数変換係数に決定することを特徴としている。

## 【 0 0 3 0 】

これによって、各入力フレームの周波数変換を行う変換係数が1つ選定されるが、入力フレーム毎に選定される変換係数は異なるため、1フレーム毎により細かく話者正規化を行うことが可能になり、いかなる音声に対してもこの音声認識

方法を適用することができる。

【0031】

また、本発明の音声認識装置は、音響特徴量と標準音韻モデルとを用いて、フレーム毎に各音韻同士の類似度あるいは距離の比を重みとして算出する重み算出部を更に備え、音声認識部が入力音声正規化する周波数変換後に算出する類似度あるいは距離に、重みを掛け合わせたものを使用して認識することを特徴としている。

【0032】

これによって、各入力フレームのスペクトルを周波数変換する変換係数は複数選定され、重み付き合算処理が行われ、かつ入力フレーム毎に重みの組値が異なる。このため、1フレーム毎により精度良く話者正規化を行うことが可能になり、いかなる音声に対してもこの音声認識装置を用いることができる。

【0033】

また、本発明の音声認識装置の変換条件判定部は、複数の前記フレームでの最尤の音韻に対する最適周波数変換係数となった頻度の最も多い周波数変換係数を当該最尤の音韻に対する入力音声最適変換係数と決定し、更に音韻毎の重みを、フレーム毎の最適周波数変換係数の重みに変換し、重みが0でない最適周波数変換係数を、当該フレームの入力音声を正規化する周波数変換係数に決定することを特徴としている。

【0034】

これによって、各入力フレームのスペクトルを周波数変換する変換係数は複数選定され、重み付き合算処理が行われ、かつ入力フレーム毎に重みの組値が異なる。このため、1フレーム毎により精度良く話者正規化を行うことが可能になり、いかなる音声に対してもこの音声認識装置を用いることができる。

【0035】

また、本発明の音声認識装置の類似度あるいは距離の比較は、少なくとも母音を使用することを特徴としている。

【0036】

これによって、周波数変換を行う対象として最も信頼性の高い母音のみの情報

から全区間の最適変換係数を決定することになるため、決定された最適変換係数の信頼度もより高いものにすることができる。

## 【 0 0 3 7 】

また、本発明の音声認識装置の類似度あるいは距離の比較は、母音のみを使用することを特徴としている。

## 【 0 0 3 8 】

これによって、音声認識性能が劣化することなく、標準音韻モデルのデータ量を少なく抑えることができる。

## 【 0 0 3 9 】

以下、本発明について図面を用いて説明する。

## 【 0 0 4 0 】

## (実施の形態 1)

図 1 は本発明の第 1 の実施の形態である話者正規化を用いた音声認識システムのハードウェアを示すブロック図である。図 1 において、マイクロフォン 1 0 1 は音声を取り込み、A/D変換器 1 0 2 が音声のアナログ信号をデジタル信号に変換する。シリアルコンバータ（以下「SCO」という。）1 0 3 はA/D変換器 1 0 2 からのシリアル信号をバスデータラインに送る。記憶装置 1 0 4 にはあらかじめ複数の話者から学習した音韻毎の標準音韻モデルの特徴量を統計処理した数値群である標準話者群発声モデル（以下「標準音韻モデル」という。）及び音声片モデルを連結して得られる単語モデルが記憶されている。

## 【 0 0 4 1 】

パラレル I/Oポート 1 0 5 はこの記憶装置 1 0 4 から標準音韻モデルあるいは単語モデルをバスクロックに同期してバスラインに入力し、あるいはディスプレイなどの出力装置 1 1 0 に音声認識結果を出力する。RAM 1 0 7 はデータ処理を実行するとき使用する一時記憶用のメモリであり、DMAコントローラ 1 0 6（以下「DMA」という。）が記憶装置 1 0 4、出力装置 1 1 0 及びRAM 1 0 7 の間の高速なデータ転送を行う。

## 【 0 0 4 2 】

ROM 1 0 8 は、処理プログラムやあらかじめ設定した後述する周波数変換係

数などのデータが書き込まれている。そして、これらSCO103、PIO105、DMA106、RAM107及びROM108はバス接続され、CPU109でコントロールされている。このCPUはデジタルシグナルプロセッサ(DSP)とすることも可能である。

#### 【0043】

また、SCO103乃至CPU109が音声認識装置100を構成している。

#### 【0044】

次に、図2を用いて、上記したようなハードウェア構成により実現される音声認識装置100の機能ブロックの構成について説明する。

#### 【0045】

特徴量抽出部201は、入力される音声データから音声を時分割して得られる音響特徴量を抽出する。入力される音声データ(SIG1)はデジタルデータであり、サンプリング周波数としては様々な値が使われている。例えば電話音声では8kHzが使われ、CDオーディオでは44.1kHzが使われている。ここでは、サンプリング周波数として10kHzを使用した。

#### 【0046】

また、音響特徴量を抽出する時分割単位として、窓長及びシフト幅は5ms程度から50ms程度の値が考えられるが、本実施の形態1においては、窓長を30msとし、シフト幅を15msとした。

#### 【0047】

この時間幅の音声データより、スペクトルを表現する音響特徴量を抽出する。スペクトルを表現する特徴量としては、LPCケプストラム係数、LPCメルケプストラム係数、ケプストラム係数抽出前にメルスケールで変換を行うメル化LPCケプストラム係数、MFCC、これらケプストラム係数の差分を取ったデルタケプストラムなど、種々のパラメータが知られているが、ここでは7次のLPCメルケプストラム係数を抽出する。

#### 【0048】

周波数変換部202は、特徴量抽出部201で得られた特徴量に対して周波数変換を施す。周波数変換方法は、線形に伸縮する方法やシフトする方法、非線形

関数で伸縮やシフトする方法などが知られているが、本実施の形態1では、(式1)で表現される1次オールパスフィルタ関数を用いた非線形伸縮を行った。

【0049】

【数1】

$$\tilde{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (1)$$

【0050】

この(式1)の $\alpha$ を周波数変換係数(以下「変換係数」という。)という。変換係数 $\alpha$ は、本来は可変の値であるが、本実施の形態1においては、処理の都合上、「-0.15」、「-0.1」、「-0.05」、「0」、「+0.05」、「+0.10」、「+0.15」の7つの離散値 $\alpha_1$ 乃至 $\alpha_7$ を用いた。これらを以下では変換係数群と呼ぶこととする。

【0051】

周波数変換部202は、設定された変換係数を用いて(式1)により周波数変換処理を行う。

【0052】

変換係数設定部203は、周波数変換部202に変換係数を設定する。音響特徴量比較部204は標準音韻モデル205から標準音韻モデルデータを読み込み、周波数変換部202から得た複数の変換後の入力音響特徴量(以下「変換後特徴量」という。)とを比較する。また、この比較結果を比較結果蓄積部206に蓄積する。

【0053】

また、標準音韻モデル205は、以下に示す24音韻についての特徴量を統計処理した数値群からなっている。このデータは記憶装置104に記憶されている。

【0054】

/a/, /o/, /u/, /i/, /e/, /j/, /w/, /m/, /n/,  
/ng/, /b/, /d/, /r/, /z/, /hv/, /hu/, /s/,



/c/、/p/、/t/、/k/、/y v/、/y u/、/n/

この音韻の選択については、電子情報通信学会論文誌D-I I NO. 12 pp. 2096-pp. 2103において述べられている。

【0055】

単語モデル210は音声片モデルを連結して得られる認識対象語辞書であり、認識対象標準音響モデルの1例に相当する。

【0056】

また、標準音韻モデル205と単語モデル210は共に記憶装置104に記憶されており、どちらも同じ標準話者群の同じ発声群を入力として学習している。

【0057】

変換条件判定部207は、比較結果蓄積部206に蓄積された結果から音声認識に使用する変換係数を決定する。

【0058】

特徴量蓄積部208は、特徴量抽出部201で抽出された特徴量を音声認識処理が完了するまで一時的に蓄積するメモリであり、RAM107の一部が相当する。

【0059】

音声認識処理部209は周波数変換された特徴量と単語モデル210との類似度あるいは距離を演算し単語を決定する。また、この認識結果を出力装置110に出力する。

【0060】

このような機能構成の音声認識装置100の動作について、図3に示す流れ図を用いて説明する。

【0061】

まず、特徴量抽出部201がマイクロフォン101から入力された音声について1フレーム毎に音響特徴量を抽出する（ステップS301）。そして、抽出した特徴量を周波数変換部202へ出力すると同時に、特徴量蓄積部208に蓄積する。

【0062】

次に、変換係数設定部 2 0 3 が所定の変換係数を周波数変換部 2 0 2 に設定する。

#### 【 0 0 6 3 】

周波数変換部 2 0 2 は音響特徴量をこの変換係数での（式 1）により変換し、変換後特徴量を求める。この変換を変換係数群すべての変換係数の場合について行う。これにより、1 フレーム毎に変換係数の数だけ変換後特徴量が算出される（ステップ S 3 0 2）。

#### 【 0 0 6 4 】

次に、音響特徴量比較部 2 0 4 は、算出された変換後特徴量から一つを選び、標準音韻モデル 2 0 5 から読み出した全ての音韻の標準音韻モデルと比較する。この比較は単一フレーム同士を照合する方法と前後の数フレームを用いて照合する方法とが考えられる。本実施の形態 1 では入力フレームの前後の各 3 フレームを加えて幅 7 フレームの入力を標準音韻モデル 2 0 5 に含まれる標準話者の標準音韻モデルと比較する（ステップ S 3 0 3）。そして、この比較処理の結果を比較結果蓄積部 2 0 6 に蓄積する。なお、音響特徴量比較部 2 0 4 は算出した変換後特徴量全てに対してこの比較処理を行う。

#### 【 0 0 6 5 】

この変換後特徴量と標準音韻モデルとの比較方法としては、標準話者群の発声モデルとして分布を持った統計処理を採用して音韻認識を行って求められる類似度を使用する方法と、標準話者群の発声モデルとして音韻別の代表値を採用して求められる物理的な距離を使用する方法などが考えられる。

#### 【 0 0 6 6 】

この話者正規化のために使用する音韻をモデル化した標準音韻モデル 2 0 5 については、2 通りの例を示している。

#### 【 0 0 6 7 】

第一の例では、標準話者群の発声モデルとして分布を持った統計処理を採用して音韻認識を行って求められる類似度を使用する場合においては、音韻認識のための類似度を求める尺度としてマハラノビス汎距離を使用しており、標準話者の発声から各音韻に相当する発声部分の連続 7 フレーム分の音響特徴量を集め、平

均値及び共分散行列を求めた後で、マハラノビス汎距離を求めるための係数ベクトルに変換した数値群で構成されている。

## 【0068】

第二の例では、標準話者群の発声モデルとして音韻別の代表値を採用して求められる物理的な距離を使用する場合であって、標準話者の発声から各音韻に相当する発声部分の連続7フレーム分の音響特徴量の平均ベクトル群から構成されている。

## 【0069】

なお、マハラノビス汎距離については、例えば、特開昭60-67996号公報に開示されている。

## 【0070】

これら音韻認識による類似度を使用する例と音韻別代表値との距離を使用する例の2つの例についての結果を後で述べる。

## 【0071】

また、比較結果蓄積部206に蓄積されるデータは、入力1フレーム毎に24音韻の数の音韻認識によって求められる類似度や音韻別代表値との距離となる。

## 【0072】

上記のステップS301乃至ステップ303を音声区間の全フレームについて行う。

## 【0073】

次に、変換条件判定部207は、各入力フレームの各音韻に対する最も高い類似度を示す変換係数を(式2)に沿って決定する(ステップS304)。

## 【0074】

【数2】

$$\hat{\alpha} = \arg \max_{\alpha} L(X^{\alpha} | \alpha, \theta) \quad (2)$$

## 【0075】

(式2)において、Lは類似度をあらわし、 $X^{\alpha}$ は(式1)に沿った周波数変換によるスペクトル、 $\alpha$ は変換係数、 $\theta$ は標準音韻モデルをあらわす。そして、

スペクトル  $X_\alpha$  と標準音韻モデル  $\theta$  の類似度を最大とする変換係数  $\alpha$  を探索し決定する。本実施の形態 1 では、処理の関係上 7 つの離散値  $\alpha_1$  乃至  $\alpha_7$  を用いているので、7 つの離散値すべてを当てはめた場合の類似度を比較し、最も高い類似度を得られる変換係数  $\alpha$  を選択し決定することになる。

【0076】

音韻特徴量比較の結果が距離である場合には、最も近い距離を示す変換係数を(式 3) に沿って決定する。

【0077】

【数 3】

$$\hat{\alpha} = \arg \min_{\alpha} D(X^\alpha | \alpha, \theta) \quad (3)$$

【0078】

(式 3) において、 $D$  は距離をあらわし、 $X_\alpha$  は(式 1) に沿った周波数変換によるスペクトル、 $\alpha$  は変換係数、 $\theta$  は標準音韻モデルをあらわす。そして、スペクトル  $X_\alpha$  と標準音韻モデル  $\theta$  の距離を最小とする変換係数  $\alpha$  を探索し決定する。本実施の形態では、7 つの離散値すべてを当てはめた場合の距離を比較し、最も小さい距離、即ち一番近い距離を得られる変換係数  $\alpha$  を選択し決定することになる。

【0079】

次に、フレーム毎に入力との類似度が最も高い、もしくは距離が最も小さい音韻を選択し、その音韻の標準音韻モデルに近づくように変換係数を求める(ステップ S305)。

【0080】

図 8 (A) はこの様子を示す全フレームについての音韻毎の変換係数を示した図である。図 8 (A) において、1 フレーム内の音韻毎に最尤の変換係数 801 が選択され、類似度あるいは距離の算出により最尤の音韻 802 が決定される。そして、その音韻に対応する変換係数 803 が求められる。例えば、ステップ 305 により 1 フレーム目の最尤の条件が、音韻が / a / であり、変換係数が  $\alpha_4$  であると選出された場合、その周波数変換に使用した変換係数  $\alpha_4$  が 1 フレーム

目の変換係数となる。

【0081】

次に、変換条件判定部207はステップS305で求めたフレーム毎に、選択された音韻に対応する最適変換係数の、音声全区間に渡る出現頻度を累積する。そして、累積された出現頻度を比較し、最も出現頻度が高かった変換係数を全区間の最適変換係数として決定し、変換係数設定部203に通知する（ステップS306）。図8（B）はこの変換係数と累積数との関係を示す図である。図8（B）では $\alpha 4$ が最も回数が多いため、 $\alpha 4$ が最適変換係数となる。

【0082】

以上のステップS301乃至ステップS306により、音声認識処理で使用する周波数変換係数が求まる。

【0083】

次に、変換係数設定部203は周波数変換部202に通知された変換係数を設定する。周波数変換部202はこれを受けて、特徴量蓄積部208から蓄積されている特徴量を読み出し、1フレーム目から全音声区間に渡って周波数変換を行う（ステップS307）。この結果である変換後特徴量を音声認識処理部209へ出力する。

【0084】

以上のステップS301乃至ステップS307が話者正規化处理である。

【0085】

次に、音声認識処理部209は得られた変換後特徴量を用いて、音声認識処理を行う。この処理方法としては、隠れマルコフモデルを用いた方法、ダイナミックタイムワーピングによる方法、ニューラルネットワークによる方法などが知られているが、本実施の形態1では、特開平4-369696号公報、特開平5-150797号公報及び特開平6-266393号公報に開示されている方法を用いた。音声認識処理部209は、入力と単語モデルとを用いて音声認識処理を行い、認識された単語を音声認識結果として出力装置110に出力する（ステップS308）。

【0086】

以上のように、本実施の形態 1 では音韻認識では十分とされる 2 4 音韻すべてについての類似度あるいは距離から最適変換係数を決定しているため、いかなる種類の発声であってもこの音声認識方法を適用することができる。

【0087】

また、本実施の形態 1 のステップ S 3 0 7 において、選択された音韻全ての最適変換係数の出現回数を累積したが、選択された音韻が母音であった場合のみ回数をカウントすることも可能である。これにより、周波数変換を行う対象として最も信頼性の高い母音のみの情報から全区間の最適変換係数を決定することになるため、決定された最適変換係数の信頼度もより高いものにすることができる。

【0088】

図 1 1 (A) は本実施の形態 1 により、話者正規化を行った場合と、行わない場合との音声認識の結果を示す。このテストは 1 0 0 単語入力について、1 0 0 単語の登録辞書を用い、不特定話者 3 名で行った。話者正規化を行うことにより認識率が 7 % 乃至 2 1 % 向上した。これより、無声音と有声音との区間検出なしの継続長固定の音素認識もしくは入力と標準音韻モデルとの距離計算であって、認識対象語辞書を使用せずに話者正規化を行っても、本願発明によれば上記効果が得られることが確認できる。

【0089】

なお、本実施の形態 1 では全音声区間に適応する変換係数は全音声区間の周波数変換処理を行った後に決定しているが、変換係数のいずれかが所定の回数最適変換係数をして選択された時点で、全音声区間に適応する変換係数とすることも可能である。これにより、音声認識時間の短縮を図ることができる。

【0090】

(実施の形態 2)

図 4 は本発明の第 2 の実施の形態である音声認識装置の機能構成を示す。

【0091】

第 1 の実施の形態とは、音響特徴量比較部 2 0 4 が周波数変換部 2 0 2 からの出力の他に特徴量抽出部 2 0 1 の出力である音響特徴量と標準音韻モデル 2 0 5 とを比較する点が異なる。更に、変換条件判定部 2 0 7 がこの音響特徴量比較部

2 0 4 の比較結果である後述する代表音韻を用いて変換条件の判定を行う点が異なる。

【 0 0 9 2 】

図 5 は本発明の第 2 の実施の形態である音声認識方法を示す。

【 0 0 9 3 】

以下に、図 4 と図 5 を用いて本実施の形態 2 の音声認識の動作を説明する。

【 0 0 9 4 】

この方法において、前半のステップ 3 0 1 乃至ステップ S 3 0 4 の処理は実施の形態 1 と同じであり、変換条件判定部 2 0 7 が各フレームにおける音韻毎の最適変換係数を決定する。

【 0 0 9 5 】

次に、変換条件判定部 2 0 7 は各音韻別にステップ S 3 0 4 で決定された最適変換係数の出現頻度を累積する（ステップ S 5 0 1）。図 9（A）はこの処理の結果作成される音韻と変換係数との出現回数の関係を示した図の一例である。

【 0 0 9 6 】

次に、変換条件判定部 2 0 7 は各音韻別に最も頻度が高かった変換係数を選定し、当該音韻の全音声区間向け変換係数として決定する（ステップ S 5 0 2）。図 9（A）においては音韻／a／の変換係数として $\alpha_4$ が選択され、音韻／e／の変換係数としては $\alpha_3$ が選択されたことを示す。

【 0 0 9 7 】

一方、音響特徴量比較部 2 0 4 は入力フレーム全区間に対し、フレーム毎に当該入力フレームを代表する音韻を決定する（ステップ S 5 0 3）。ここでは、音響特徴量比較部 2 0 4 が特徴量抽出部 2 0 1 の出力と標準音韻モデル 2 0 5 の各音韻標準音韻モデルとを比較し、類似度が最も高いもの、あるいは音韻別代表値との距離が最も小さい音韻を代表音韻として選定する。

【 0 0 9 8 】

次に、変換条件判定部 2 0 7 は当該入力フレームの代表音韻に対応する変換係数をステップ S 4 0 2 での決定に基づいて選択する。この処理を入力フレーム全区間について行い、変換係数設定部 2 0 3 へ通知する（ステップ S 5 0 4）。図

9 (B) は全フレームの代表音韻とそれに対応する変換係数との関係を示す図の一例である。

【0 0 9 9】

次に、変換係数設定部 2 0 3 は入力フレーム毎に、適応する通知された変換係数を周波数変換部 2 0 2 に設定する。周波数変換部 2 0 2 はこれを受けて、特徴量蓄積部 2 0 8 から蓄積されている特徴量を読み出し、再度周波数変換処理を行う (ステップ S 5 0 5)。そして、この処理を全音声区間について行う。

【0 1 0 0】

以上のステップ S 3 0 1 乃至ステップ S 5 0 5 が本実施の形態 2 の話者正規化処理である。

【0 1 0 1】

次に行う音声認識処理ステップ S 3 0 8 は実施の形態 1 と同一である。

【0 1 0 2】

以上のように、本実施の形態 2 では各入力フレームの周波数変換を行う変換係数は 1 つ選定されるが、入力フレーム毎に選定される変換係数は異なるため、1 フレーム毎により細かく話者正規化を行うことが可能になり、いかなる音声に対してもこの音声認識方法を適用することができる。

【0 1 0 3】

図 1 1 (B) は本実施の形態 2 により、話者正規化を行った場合と、行わない場合との音声認識の結果を示す。このテストは 1 0 0 単語入力について、1 0 0 単語の登録辞書を用いて不特定話者 9 名で行った。話者正規化を行うことにより認識率が成人に比べ低かった子供の認識率が 8. 2 % 向上した。これより、無声音と有声音との区間検出なしの継続長固定の音素認識もしくは入力と音韻標準音韻モデルとの距離計算であって、認識対象語辞書をせずに話者正規化を行っても、本願発明によれば上記効果が得られることが確認できる。

【0 1 0 4】

(実施の形態 3)

図 6 は本発明の第 3 の実施の形態である音声認識装置の機能構成を示す。

【0 1 0 5】



第 2 の実施の形態とは、特徴量から音韻毎の重みを算出する重み算出部 6 0 1 を備えている点が異なる。

【0 1 0 6】

図 7 は本発明の第 3 の実施の形態の音声認識方法を示す。

【0 1 0 7】

以下に、図 6 と図 7 を用いて本実施の形態 3 の音声認識の動作を説明する。

【0 1 0 8】

この方法において、前半のステップ 3 0 1 乃至ステップ S 5 0 2 の処理は第 2 の実施の形態と同じであり、変換条件判定部 2 0 7 が音韻毎の最適変換係数を決定する。

【0 1 0 9】

次に、音響特徴量比較部 2 0 4 は入力音声全区間に対し、フレーム毎に音韻重みを決定する（ステップ S 7 0 1）。この重みの決定のために、まず音響特徴量比較部 2 0 4 が特徴量抽出部 2 0 1 の出力と標準音韻モデル 2 0 5 の各音韻標準音韻モデルとの類似度、あるいは音韻別代表値との距離を算出する。更に音響特徴量比較部 2 0 4 は（式 4）を用いて正規化された重みを求める。

【0 1 1 0】

（式 4）において、 $w_{ik}$  は重み、 $X$  は入力スペクトル、 $V$  は音韻別代表値ベクトル、 $k$  は音韻の種類、 $p$  は内挿の平滑度を表すパラメータであり、 $d(X, V)$  は（式 5）により求まる入力スペクトルと音韻別代表値との距離を示す。

【0 1 1 1】

【数 4】

$$w_{ik} = \frac{d(X_i, V_k)^{-p}}{\sum_k \{d(X_i, V_k)^{-p}\}} \quad (4)$$

【0 1 1 2】

【数 5】

$$d(X, V) = \|X - V\|^2 \quad (5)$$

## 【0 1 1 3】

音響特徴量比較部 2 0 4 は全音声区間について上記処理を行い、各フレームの音韻毎の重みを算出する。この算出結果として、図 1 0 (A) に示すような各フレームの音韻と、音韻毎の重みの関係が得られる。そして、この結果は比較結果蓄積部 2 0 6 に記録される。

## 【0 1 1 4】

次に、音韻重み算出部 6 0 1 がステップ S 5 0 2 で求めた全音声区間に渡っての各音韻とそれに対応する最適変換係数との関係 (図 8 (A)) と、ステップ S 7 0 1 で求めた各フレームの音韻と、音韻毎の重みの関係 (図 1 0 (A)) とから各フレームの変換係数毎の重みを算出する (ステップ S 7 0 2)。図 1 0 (B) はこの関係を示している。そして、音韻重み算出部 6 0 1 はこの算出結果を比較結果蓄積部 2 0 6 に記憶する。

## 【0 1 1 5】

次に、変換条件判定部 2 0 7 はこの各フレームの変換係数毎の重みを比較結果蓄積部 2 0 6 から読み出し、重みが '0' 以外の変換係数を変換係数設定部 2 0 3 にフレーム毎に通知する。変換係数設定部 2 0 3 は通知を受けた変換係数を周波数変換部 2 0 2 に設定する。周波数変換部 2 0 2 は当該変換係数で再度 1 フレーム目から周波数変換を行い、音響特徴量比較部 2 0 4 に変換後特徴量を出力する (ステップ S 7 0 3)。

## 【0 1 1 6】

次に、音声認識処理部 2 0 9 は比較結果蓄積部 2 0 6 から各フレームの変換係数と重みの関係を読み出し、この変換係数に対応する重みを、ステップ S 7 0 4 で得られた変換係数に掛け算する。この処理を順次変換条件判定部 2 0 7 から通知される変換係数全てについて行い合算する (ステップ S 7 0 4)。

## 【0 1 1 7】

この算出は (式 6) により行うことができる。

## 【0 1 1 8】

【数 6】

$$\tilde{X}_i = \sum_k (w_{ik} * \hat{X}_i(\hat{\alpha}_k)) \quad (6)$$

但し、 $\hat{X}_i$  は入力音声の特徴量、 $\tilde{X}_i$  は変換後特徴量、 $\hat{\alpha}_k$  は変換係数、

$w_{ik}$  は重み

【0119】

以上のステップ S301 乃至ステップ S704 が話者正規化処理である。

【0120】

次に行う音声認識処理ステップ S308 は実施の形態 1 と同一である。

【0121】

以上のステップ S703 乃至ステップ S308 の処理を全音声区間について行う。

【0122】

以上のように、本実施の形態 3 では各入力フレームのスペクトルを周波数変換する変換係数は複数選定され、重み付き合算処理が行われ、かつ入力フレーム毎に重みの組値が異なる。このため、1 フレーム毎により精度良く話者正規化を行うことが可能になり、いかなる音声に対してもこの音声認識方法を適用することができる。

【0123】

また、重みは周波数変換前の特徴量を使用して求めるため、周波数変換の影響が 2 重に周波数変換時に及ぶことを防ぐことが出来、周波数変換が悪い方向で作用する話者音声に対しても、影響を低く抑えることができる。

【0124】

図 11 (C) は本実施の形態 3 により、話者正規化を行った場合と、行わない場合との音声認識の結果を示す。このテストは 100 単語入力について、100 単語の登録辞書を用い、不特定話者 9 名で行った。話者正規化を行うことにより認識率が成人に比べ低かった子供の認識率が 9.2% 向上した。

【 0 1 2 5 】

これより、無声音と有声音との区間検出なしの継続長固定の音素認識もしくは入力と標準音韻モデルとの距離計算であって、認識対象語辞書を使用せずに話者正規化を行っても、本願発明によれば上記効果が得られることが確認できる。

【 0 1 2 6 】

なお、以上の説明では、代表的な条件で構成した例で説明したが、取りうる別な方法として列記したものを使用しても同様に実施可能である。

【 0 1 2 7 】

また、本実施例においては、話者正規化による効果を単語認識の場合について説明したが、文章の認識や会話音声の認識についても、同様に実施可能である。

【 0 1 2 8 】

【発明の効果】

本発明により、入力音声を標準話者の音響特徴量に合わせるように正規化することによって入力音声の話者による差が正規化され、認識対象語辞書を使用せずに話者正規化を行うことで、認識性能を向上させることができる。

【図面の簡単な説明】

【図 1】

本発明の実施の形態 1 による音声認識システムのハードウェアを示すブロック図

【図 2】

本発明の実施の形態 1 による音声認識装置の機能構成を示す機能ブロック図

【図 3】

本発明の実施の形態 1 による音声認識装置の処理を示す流れ図

【図 4】

本発明の実施の形態 2 による音声認識装置の機能構成を示す機能ブロック図

【図 5】

本発明の実施の形態 2 による音声認識装置の処理を示す流れ図

【図 6】

本発明の実施の形態 3 による音声認識装置の機能構成を示す機能ブロック図

【図 7】

本発明の実施の形態 3 による音声認識装置の処理を示す流れ図

【図 8】

(A) 本発明の実施の形態 1 による各フレームの音韻と変換係数の関係図

(B) 本発明の実施の形態 1 による変換係数と頻度の関係図

【図 9】

(A) 本発明の実施の形態 2 による音韻と変換係数との関係図

(B) 本発明の実施の形態 2 による各フレームの代表音韻と変換係数の関係図

【図 1 0】

(A) 本発明の実施の形態 3 による各フレームの音韻と重みの関係図

(B) 本発明の実施の形態 3 による各フレームの変換係数と重みの関係図

【図 1 1】

(A) 本発明の実施の形態 1 による音声認識の結果を示す図

(B) 本発明の実施の形態 1 による音声認識の結果を示す図

(C) 本発明の実施の形態 1 による音声認識の結果を示す図

【図 1 2】

従来の音声認識装置の処理を示す流れ図

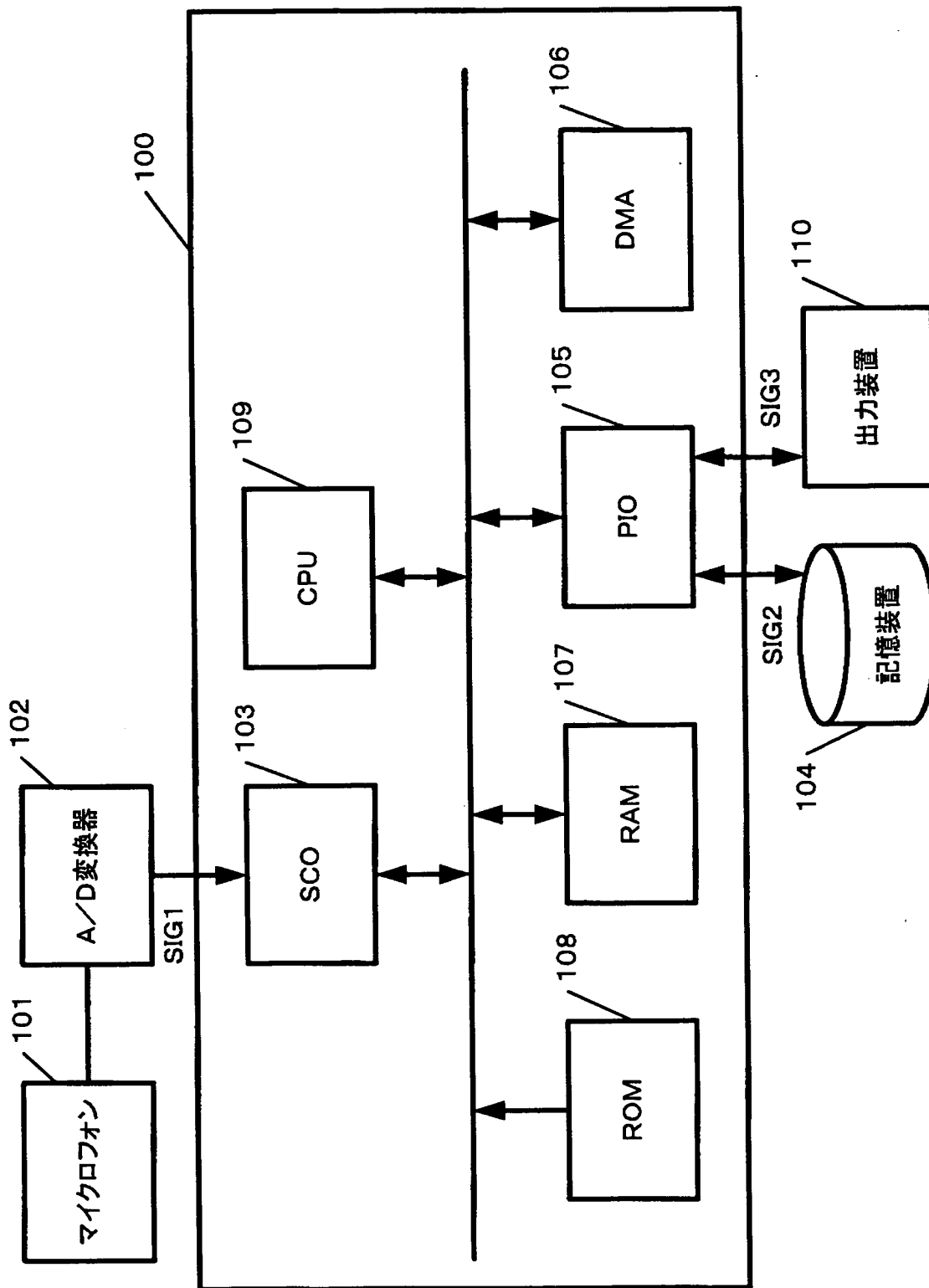
【符号の説明】

- 1 0 0 音声認識装置
- 1 0 1 マイクロフォン
- 1 0 2 A/D変換器
- 1 0 3 S C O
- 1 0 4 記憶装置
- 1 0 5 P I O
- 1 0 6 D M A
- 1 0 7 R A M
- 1 0 8 R O M
- 1 0 9 C P U
- 1 1 0 出力装置

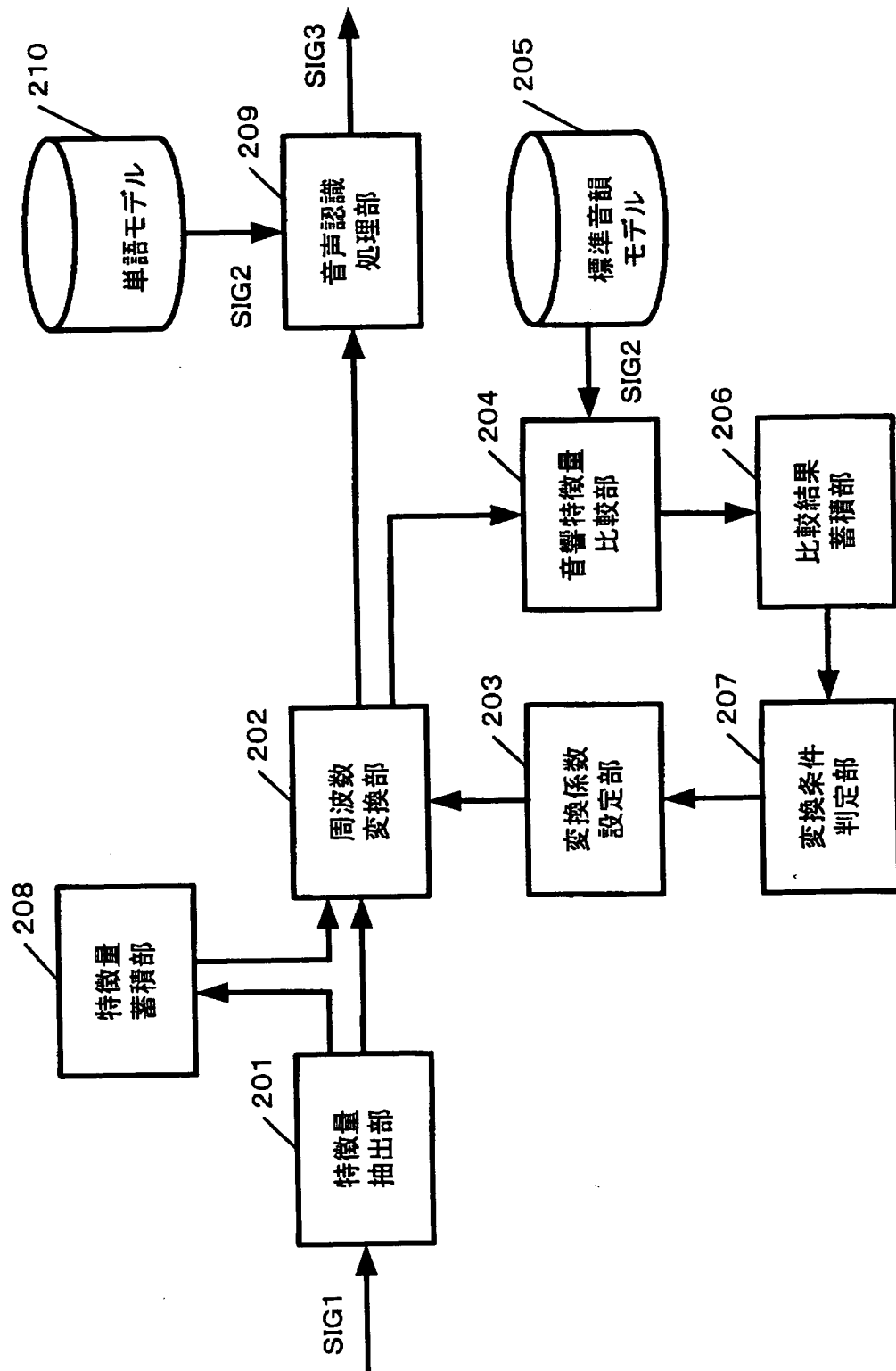
- 2 0 1 特徴量抽出部
- 2 0 2 周波数変換部
- 2 0 3 変換係数設定部
- 2 0 4 音韻特徴量比較部
- 2 0 5 標準音韻モデル
- 2 0 6 比較結果蓄積部
- 2 0 7 変換条件判定部
- 2 0 8 特徴量蓄積部
- 2 0 9 音声認識処理部
- 2 1 0 単語モデル
- 6 0 1 重み算出部

【書類名】 図面

【図 1】

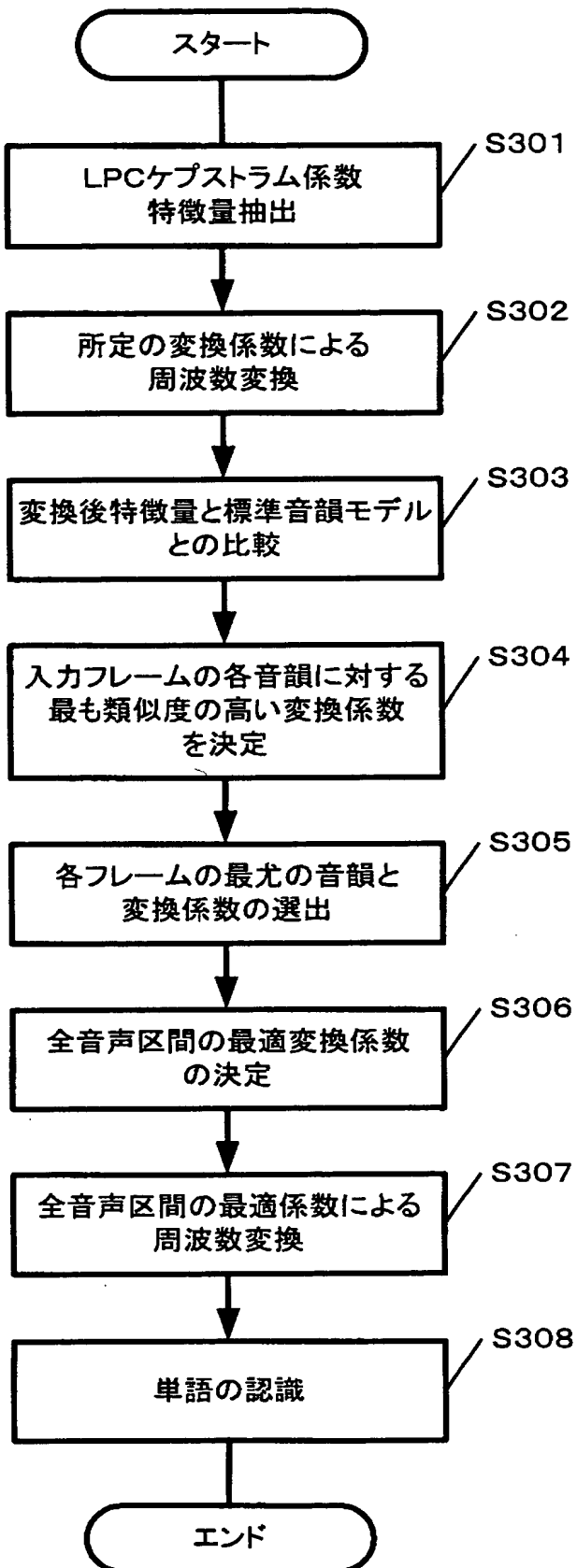


【図 2】

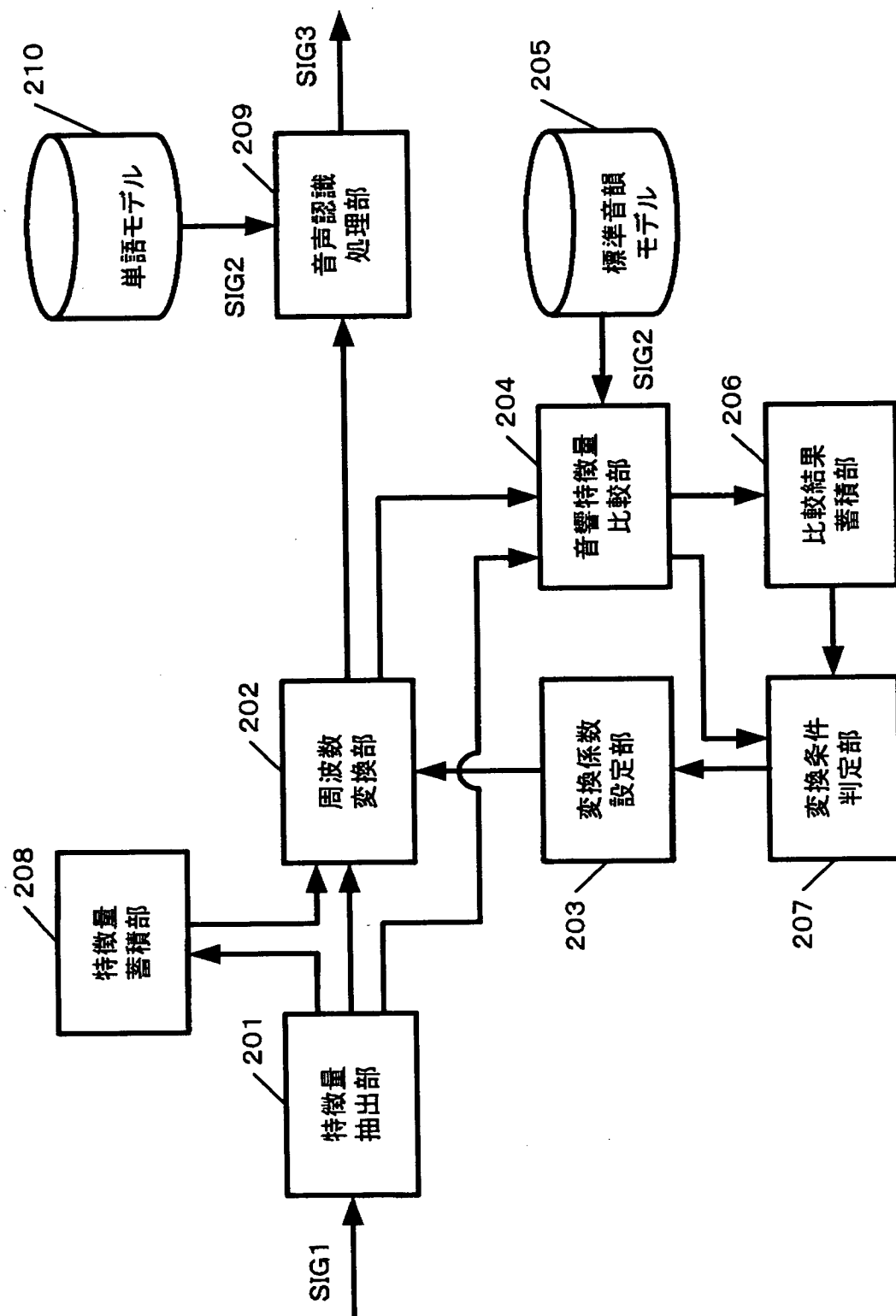




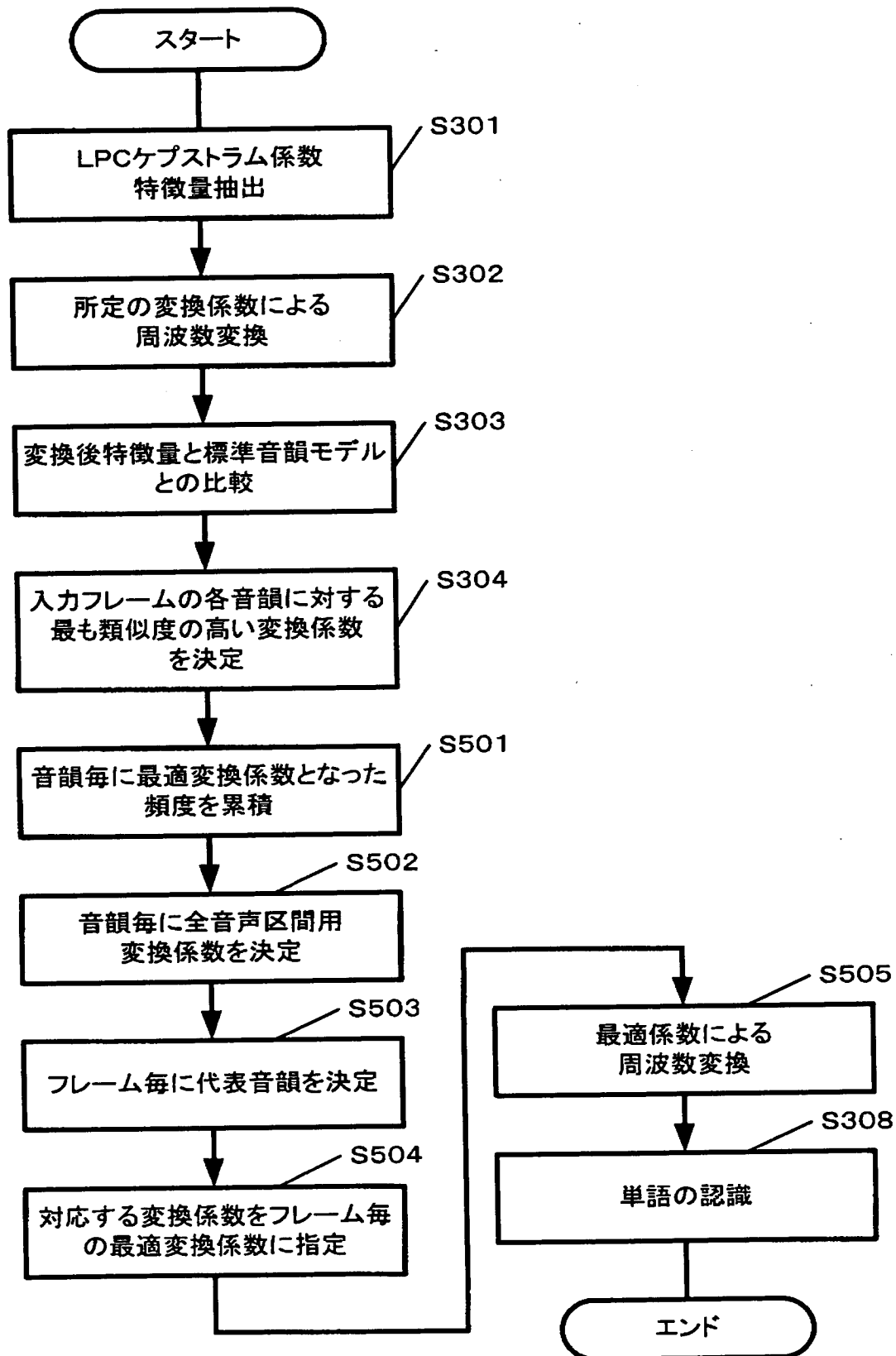
【図 3】



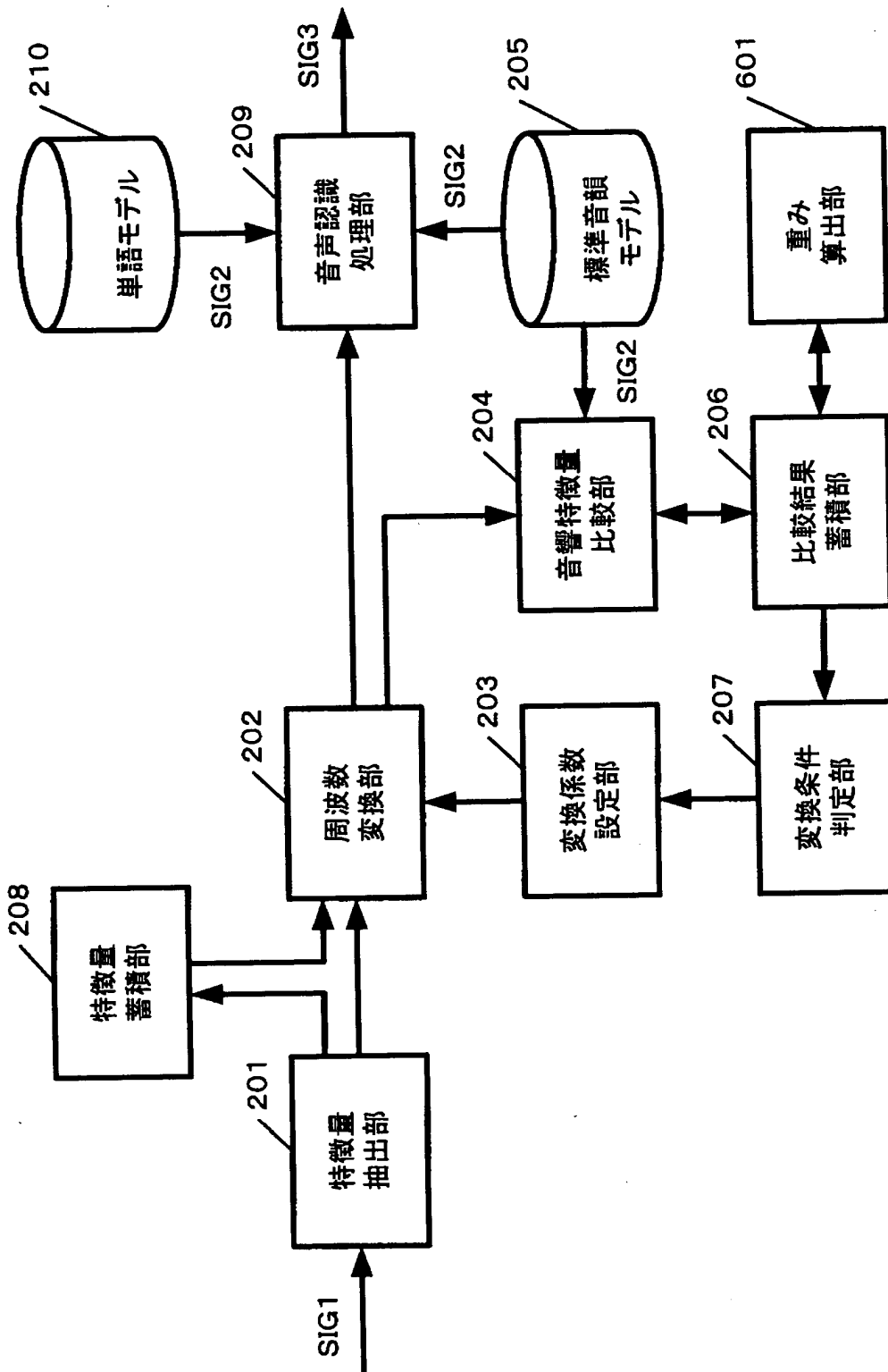
【図 4】



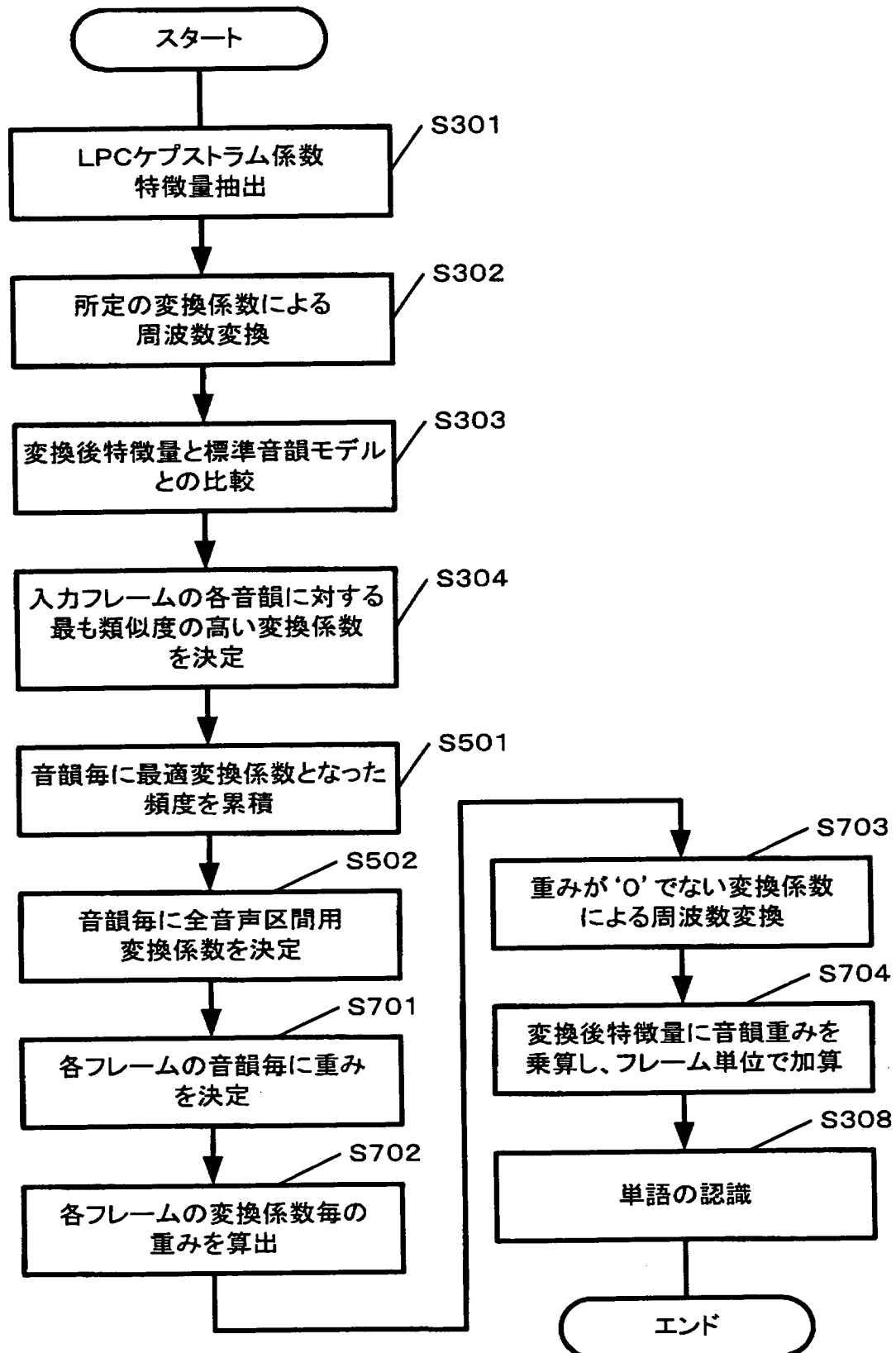
【図 5】



【図 6】



【図 7】



【図 8】

(A)

フレーム番号	801					802	803
	/a/	/e/	/i/	/n/	.....	最尤の音韻	対応する 変換係数
1フレーム目	$\alpha_4$	$\alpha_3$	$\alpha_5$			/a/	$\alpha_4$
2フレーム目	$\alpha_2$	$\alpha_2$	$\alpha_3$			/i/	$\alpha_3$
.				.		.	.
.				.		.	.
.				.		.	.

(B)

変換係数	回数
$\alpha_1$	0
$\alpha_2$	5
$\alpha_3$	48
$\alpha_4$	112
$\alpha_5$	21
$\alpha_6$	3
$\alpha_7$	0

【図 9】

(A)

音韻	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$ (単位 回数)	$\alpha_7$	累積数が最大となる 変換係数
/a/	0	0	8	26	12	0	0	$\alpha_4$
/e/	0	3	39	4	0	0	0	$\alpha_3$
.				.				.
.				.				.
.				.				.

(B)

フレーム番号	代表音韻	対応する変換係数
1フレーム目	/a/	$\alpha_4$
2フレーム目	/k/	$\alpha_4$
.	.	.
.	.	.
iフレーム目	/n/	$\alpha_3$
.	.	.
.	.	.

【図 10】

(A)

	/a/	/e/	/i/	/n/	.....
177-△目	$w_{1a}$	$w_{1e}$	$w_{1i}$	$w_{1n}$	.....
277-△目	$w_{2a}$	$w_{2e}$	$w_{2i}$	$w_{2n}$	.....
.				.	
.				.	
.				.	

(B)

	$\alpha_1$	$\alpha_2$	$\alpha_3$	.....
177-△目	$w_{1a} + w_{1u}$	$w_{1i}$	$w_{1p} + w_{1e} +$	.....
277-△目	$w_{2k} + w_{2u}$	$w_{2a}$	$w_{2e} + w_{2r} +$	.....
.			.	
.			.	
.			.	



【図 1 1】

(A)

話者	話者正規化あり (%)	話者正規化なし (%)
話者 A (男児)	8 9	8 2
話者 B (男児)	8 3	7 5
話者 C (女児)	9 6	7 5

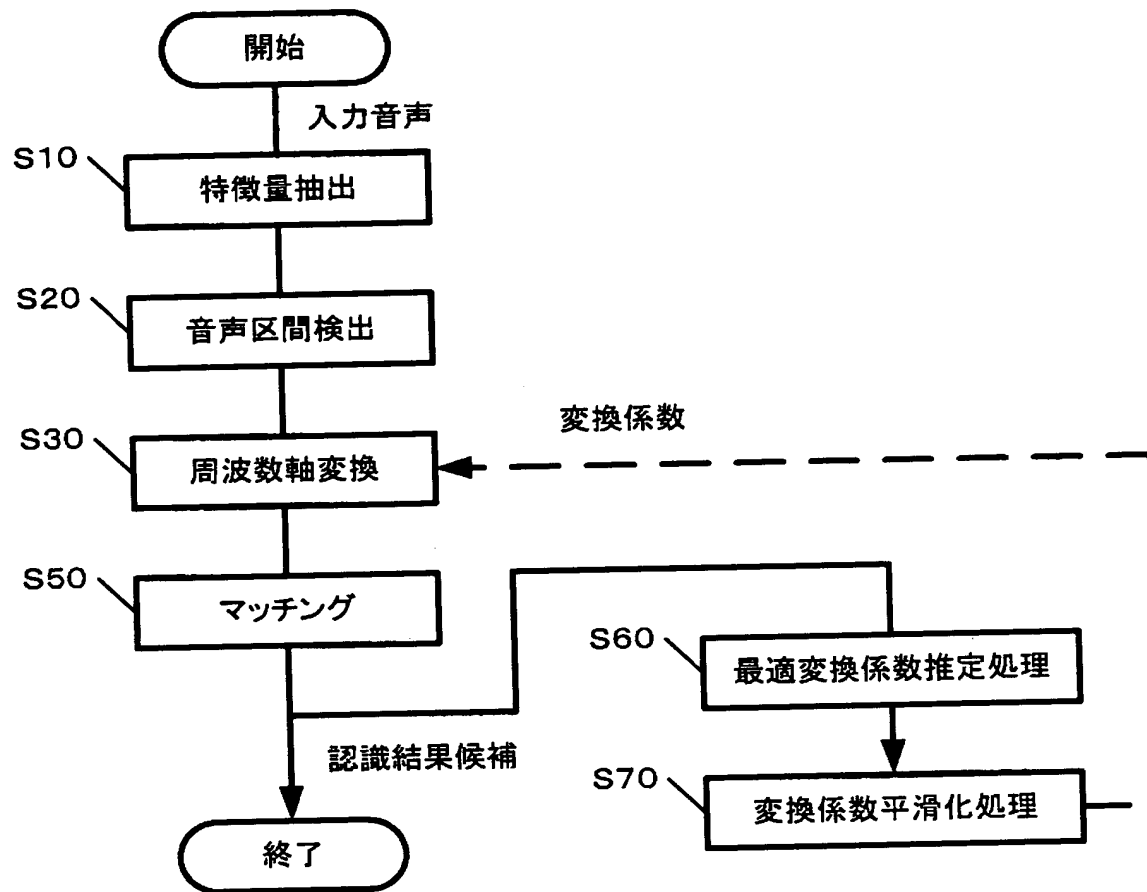
(B)

話者	話者正規化あり (%)	話者正規化なし (%)
成人男性 (2 名平均)	9 9	9 9
成人女子 (3 名平均)	9 9 . 3	9 8 . 7
子供 (4 名平均)	9 2	8 3 . 8

(C)

話者	話者正規化あり (%)	話者正規化なし (%)
話者 A (男児)	9 9	9 9
話者 B (男児)	9 9 . 3	9 8 . 7
話者 C (女児)	9 3	8 3 . 8

【図12】



【書類名】 要約書

【要約】

【課題】 音声を発する話者が入れ替わっても、入力音声の個人差を補正して音声認識性能を向上させることを目的とする。

【解決手段】 入力音声の音響特徴量を抽出する特徴量抽出部 2 0 1 と、所定の周波数変換係数を用いて周波数変換する周波数変換部 2 0 2 と、変換後特徴量と標準音韻モデルの音響特徴量を用いて類似度あるいは距離を比較して、最尤の音韻と最適周波数軸変換係数とを算出する音響特徴量比較部 2 0 4 と、これらを用いて入力音声を正規化する周波数変換係数を選択する変換条件判定部 2 0 7 と、周波数変換部 2 0 2 からの変換後特徴量を入力とする音声認識処理部 2 0 9 とを備え、音声認識処理部 2 0 9 が選択された周波数変換係数を用いて再度周波数変換された変換後特徴量を用いて認識を行うことにより、音声認識性能を向上させることができる。

【選択図】 図 2

出 願 人 履 歴 情 報

識別番号 [ 0 0 0 0 0 5 8 2 1 ]

1. 変更年月日	1 9 9 0 年 8 月 2 8 日
[変更理由]	新規登録
住 所	大阪府門真市大字門真 1 0 0 6 番地
氏 名	松下電器産業株式会社